Bayesian sparsity enforcing methods for general inverse problems

Ali Mohammad-Djafari
Laboratoire des Signaux et Systèmes,
UMR8506 CNRS-SUPELEC-UNIV PARIS SUD 11
SUPELEC, 91192 Gif-sur-Yvette, France
http://lss.supelec.free.fr

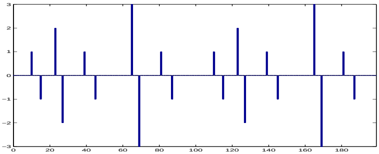
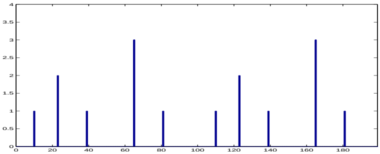
Email: djafari@lss.supelec.fr
http://djafari.free.fr

Seminar at Sharif University, EECS Dep. December 16, 2014

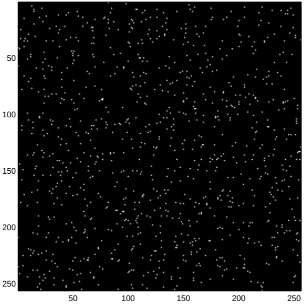
Contents

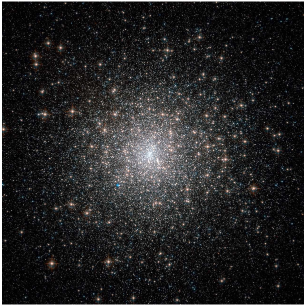
1. Sparse signals and images

- Sparse signals: Direct sparsity

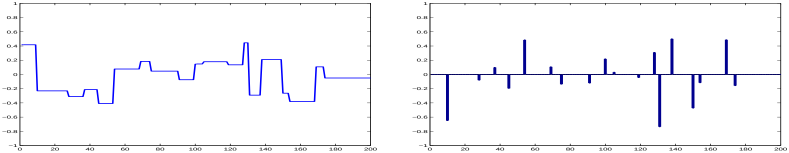
- Sparse images: Direct sparsity



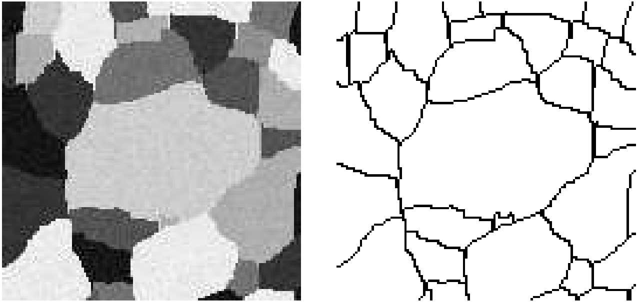


Sparse signals and images

- Sparse signals in a Transform domain



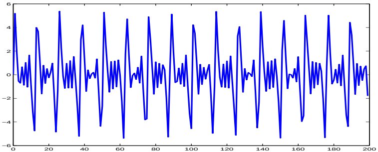
- Sparse images in a Transform domain



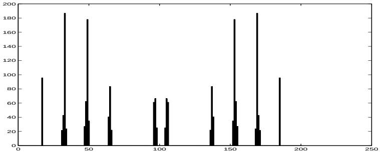
Sparse signals and images

- Sparse signals in Fourier domain

Time domain



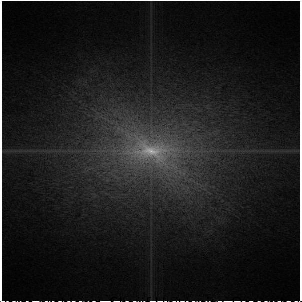
Fourier domain



- Sparse images in wavelet domain

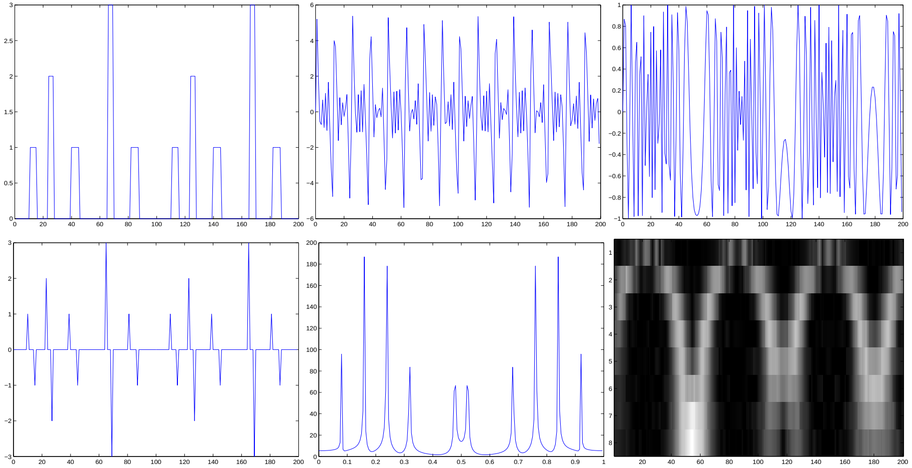
Space domain

Fourier domain

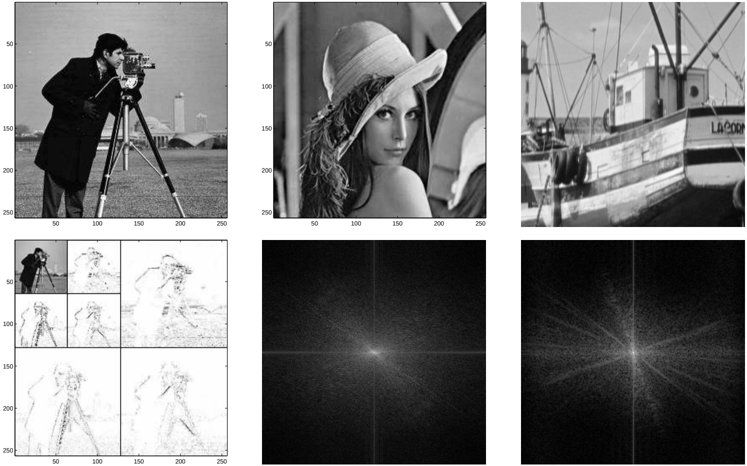


Sparse signals and images

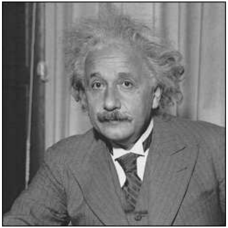
- Sparse signals: Sparsity in a Transform domaine



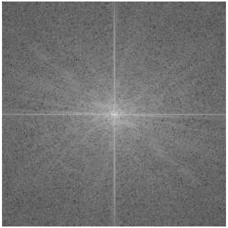
Sparse signals and images



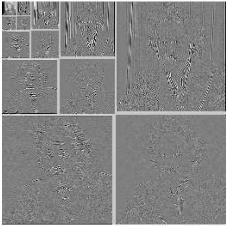
Sparse signals and images (Fourier and Wavelets domain)



Image



Fourier



Wavelets

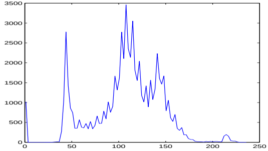
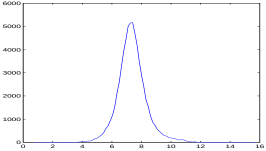
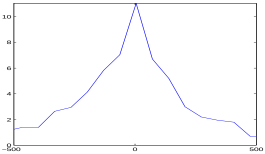


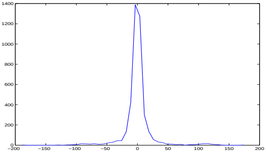
Image hist.



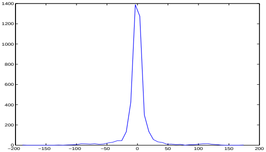
Fourier coeff. hist.



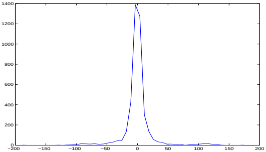
Wavelet coeff. hist.



bands 1-3



bands 4-6



bands 7-9

2. First ideas: some history

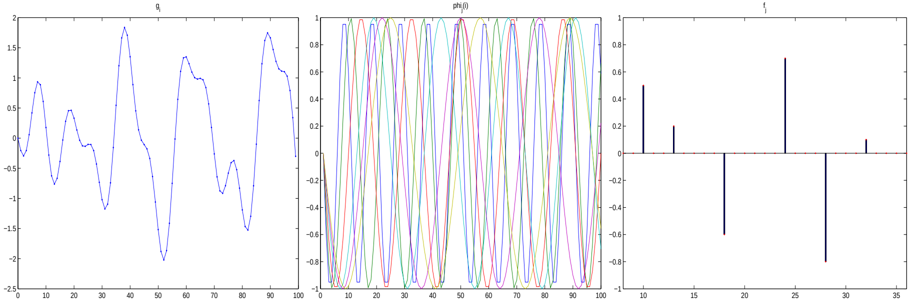
- 1948: Shannon:
  Sampling theorem and reconstruction of a band limited signal
- 1993-2007:
  - Mallat, Zhang, Candès, Romberg, Tao and Baraniuk:
    Non linear sampling, Compression and reconstruction,
  - Fuch: Sparse representation
  - Donoho, Elad, Tibshirani, Tropp, Duarte, Laska:
    Compressive Sampling, Compressive Sensing
- 2007-2012:
  Algorithms for sparse representation and compressive Sampling: Matching Pursuit (MP), Projection Pursuit Regression, Pure Greedy Algorithm, OMP, Basis Poursuit (BP), Dantzig Selector (DS), Least Absolute Shrinkage and Selection Operator (LASSO), Iterative Hard Thresholding...
- 2003-2012:
  Bayesian approach to sparse modeling
  Tipping, Bishop: Sparse Bayesian Learning,
  Relevance Vector Machine (RVM), ...

3. Modeling and representation

- Modeling via a basis
  (codebook, overcomplete dictionnary, Design Matrix)

$$g(t) = \sum_{j=1}^{N} f_j \, \phi_j(t), \; t = 1, \cdots, T \longrightarrow \boldsymbol{g} = \boldsymbol{\Phi}' \boldsymbol{f}$$

$g(t)$ $\phi_j(t)$ f_j



$T = 100$ $N = 35$

3. Modeling and representation

- Modeling via a basis
  (codebook, overcomplete dictionnary, Design Matrix)

$$g(t)=\sum_{j=1}^{N} f_j\,\phi_j(t),\; t=1,\cdots,T \longrightarrow \boldsymbol{g}=\boldsymbol{\Phi}'\,\boldsymbol{f}$$

- When $T \geq N$

$$\widehat{f}_j=\arg\min_{f_j}\left\{\sum_{t=1}^{T}\left|g(t)-\sum_{j=1}^{N} f_j\,\phi_j(t)\right|^2\right\}\longrightarrow$$

$$\widehat{\boldsymbol{f}}=\arg\min_{\boldsymbol{f}}\left\{\|\boldsymbol{g}-\boldsymbol{\Phi}'\boldsymbol{f}\|^2\right\}=[\boldsymbol{\Phi}\boldsymbol{\Phi}']^{-1}\boldsymbol{\Phi}\boldsymbol{g}$$

- When orthogonal basis: $\boldsymbol{\Phi}\boldsymbol{\Phi}'=\boldsymbol{I}\longrightarrow\widehat{\boldsymbol{f}}=\boldsymbol{\Phi}\boldsymbol{g}$

$$\widehat{f}_j=\sum_{t=1}^{N} g(t)\,\phi_j(t) =< g(t),\phi_j(t) >$$

- Application in Compression, Transmission and Decompression

Modeling and representation

- When overcomplete basis $N > T$: Infinite number of solutions for $\boldsymbol{\Phi}' \boldsymbol{f} = \boldsymbol{g}$. We have to select one:

$$\widehat{\boldsymbol{f}} = \arg \min_{\boldsymbol{f}:\ \boldsymbol{\Phi}'\boldsymbol{f}=\boldsymbol{g}} \left\{ \|\boldsymbol{f}\|_2^2 \right\}$$

  or writing differently:

$$\text{minimize} \quad \|\boldsymbol{f}\|_2^2 \quad \text{subject to} \quad \boldsymbol{\Phi}'\boldsymbol{f} = \boldsymbol{g}$$

  resulting to:

$$\widehat{\boldsymbol{f}} = \boldsymbol{\Phi}[\boldsymbol{\Phi}'\boldsymbol{\Phi}]^{-1}\boldsymbol{g}$$

- Again if $\boldsymbol{\Phi}'\boldsymbol{\Phi} = \boldsymbol{I} \longrightarrow \widehat{\boldsymbol{f}} = \boldsymbol{\Phi}\boldsymbol{g}$.
- No real interest if we have to keep all the N coefficients:
- Sparsity:

$$\text{minimize} \quad \|\boldsymbol{f}\|_0 \quad \text{subject to} \quad \boldsymbol{\Phi}'\boldsymbol{f} = \boldsymbol{g}$$

  or

$$\text{minimize} \quad \|\boldsymbol{f}\|_1 \quad \text{subject to} \quad \boldsymbol{\Phi}'\boldsymbol{f} = \boldsymbol{g}$$

Sparse decomposition

- Strict sparsity and exact reconstruction

$$\text{minimize} \quad \|\boldsymbol{f}\|_0 \quad \text{subject to} \quad \boldsymbol{\Phi}'\boldsymbol{f} = \boldsymbol{g}$$

$\|\boldsymbol{f}\|_0$ is the number of non-zero elements of $\boldsymbol{f}$

  - Matching Pursuit (MP) [Mallat & Zhang, 1993]
  - Orthogonal Matching Pursuit (OMP) [Lin, Huang et al., 1993]
  - Projection Pursuit Regression
  - Greedy Algorithms
  - Iterative Hard Thresholding (IHT) [Marvasti et al]

- Sparsity enforcing and exact reconstruction

$$\text{minimize} \quad \|\boldsymbol{f}\|_1 \quad \text{subject to} \quad \boldsymbol{\Phi}'\boldsymbol{f} = \boldsymbol{g}$$

  - Basis Pursuit (BP)
  - Block Coordinate Relaxation

Sparse decomposition

- Strict sparsity and approximate reconstruction

$$\text{minimize} \quad \|\boldsymbol{f}\|_0 \quad \text{subject to} \quad \|\boldsymbol{g} - \boldsymbol{\Phi}'\boldsymbol{f}\|^2 < c$$

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \left\{\|\boldsymbol{f}\|_0 + \mu\|\boldsymbol{g} - \boldsymbol{\Phi}'\boldsymbol{f}\|^2\right\} = \arg\min_{\boldsymbol{f}} \left\{\|\boldsymbol{g} - \boldsymbol{\Phi}'\boldsymbol{f}\|^2 + \lambda\|\boldsymbol{f}\|_0\right\}$$

- Sparsity enforcing and approximate reconstruction

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \left\{\|\boldsymbol{g} - \boldsymbol{\Phi}'\boldsymbol{f}\|^2 + \lambda\|\boldsymbol{f}\|_1\right\}$$

$$J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{\Phi}'\boldsymbol{f}\|^2 + \lambda\|\boldsymbol{f}\|_1 = \|\boldsymbol{g} - \boldsymbol{\Phi}'\boldsymbol{f}\|^2 + \lambda\sum_j |f_j|$$

  - Main Algorithm: LASSO [Tibshirani 2003]

$$\text{minimize} \quad \|\boldsymbol{g} - \boldsymbol{\Phi}'\boldsymbol{f}\|^2 \quad \text{subject to} \quad \|\boldsymbol{f}\|_1 < \tau$$

Sparse Decomposition Algorithms

- LASSO:

$$J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{\Phi}'\boldsymbol{f}\|^2 + \lambda \sum_j |f_j|$$

- Other Criteria
  - L_p

$$J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{\Phi}'\boldsymbol{f}\|^2 + \lambda_1 \sum_j |f_j|^p, \quad 1 < p \leq 2$$

  - Elastic net

$$J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{\Phi}'\boldsymbol{f}\|^2 + \sum_j \left(\lambda_1 |f_j| + \lambda_2 |f_j|^2\right)$$

  - Group LASSO

$$J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{\Phi}'\boldsymbol{f}\|^2 + \lambda_1 \sum_j |f_j| + \lambda_2 \sum_j |f_j - f_{j-1}|^2$$

  - Weighted L1:

$$J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{\Phi}'\boldsymbol{f}\|^2 + \lambda \sum_j |w_j f_j|$$

λ

Multi Dimensional signals: PCA, SPCA, BSS, ...

$$g_i(t) = \sum_{j=1}^{N} \boldsymbol{\Phi}_{ij} f_j(t),\ i = 1, \cdots, M,\ t = 1, \cdots, T$$

$$\boldsymbol{g}(t) = \boldsymbol{\Phi}\, \boldsymbol{f}(t),\ t = 1, \cdots, T$$

$$\boldsymbol{G} = \boldsymbol{\Phi}\, \boldsymbol{F}, \quad \text{with } \boldsymbol{G}\ [M \times T], \quad \boldsymbol{\Phi}\ [M \times N], \quad \boldsymbol{F}\ [N \times T]$$

- $f_j(t)$ factors, sources, codes
- $\boldsymbol{\Phi}$ Loading matrix (Factor Analysis),
  Mixing matrix (Blind Sources Separation),
  Design matrix (Sparse coding, Compressed Sensing)
- Objective: Find $\boldsymbol{\Phi}$ and $f_j(t)$

$$J(\boldsymbol{f}(t), \boldsymbol{\Phi}) = \sum_t \|\boldsymbol{g}(t) - \boldsymbol{\Phi} \boldsymbol{f}(t)\|^2 + \lambda_1 \sum_i \sum_j |\boldsymbol{\Phi}_{ij}| + \lambda_2 \sum_t \sum_j |f_j(t)|$$

$$J(\boldsymbol{F}, \boldsymbol{\Phi}) = \|\boldsymbol{G} - \boldsymbol{\Phi} \boldsymbol{F}\|^2 + \lambda_1 \|\boldsymbol{\Phi}\|_1 + \lambda_2 \|\boldsymbol{F}\|_1$$

Matrix Decomposition or Approximation

- Matrix approximation:
  Find an approximate matrix $\widehat{\boldsymbol{G}} = \boldsymbol{\Phi F}$ for $\boldsymbol{G}$ with some degrees of sparsity in the elements of $\boldsymbol{\Phi}$ and $\boldsymbol{F}$.

$$J(\boldsymbol{F}, \boldsymbol{\Phi}) = \|\boldsymbol{G} - \boldsymbol{\Phi F}\|^2 + \lambda_1 \|\boldsymbol{\Phi}\|_1 + \lambda_2 \|\boldsymbol{F}\|_1$$

- Low rank Matrice decomposition:

$$\widehat{\boldsymbol{G}} = \sum_{k=1}^{K} d_k \boldsymbol{u}_k \boldsymbol{v}_k' = \boldsymbol{U D V}$$

with some degrees of sparsity in the elements of $\boldsymbol{u}$ and $\boldsymbol{v}$.

$$J(\boldsymbol{U}, \boldsymbol{V}) = \|\boldsymbol{G} - \boldsymbol{U D V}\|^2 + \lambda_1 \|\boldsymbol{U}\|_1 + \lambda_2 \|\boldsymbol{V}\|_1$$

4. Sparse decomposition: Bayesian MAP interpretation

- Sparsity enforcing and approximate reconstruction

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \left\{ \|\boldsymbol{g} - \boldsymbol{\Phi}'\boldsymbol{f}\|^2 + \lambda \|\boldsymbol{f}\|_1 \right\}$$

- Equivalent to MAP estimate in a Bayesian approach
  - Bayesian approach: $\boldsymbol{g} = \boldsymbol{\Phi}'\boldsymbol{f} + \boldsymbol{\epsilon}$

$$p(\boldsymbol{f}|\boldsymbol{g}) = \frac{p(\boldsymbol{g}|\boldsymbol{f})\, p(\boldsymbol{f})}{p(\boldsymbol{g})} \propto p(\boldsymbol{g}|\boldsymbol{f})\, p(\boldsymbol{f})$$

$$\begin{cases} p(\boldsymbol{g}|\boldsymbol{f}) = \mathcal{N}(\boldsymbol{g}|\boldsymbol{\Phi}'\boldsymbol{f}, \sigma_\epsilon^2) \propto \exp\left[\frac{-1}{2\sigma_\epsilon^2} \|\boldsymbol{g} - \boldsymbol{\Phi}'\boldsymbol{f}\|^2\right] \\ p(\boldsymbol{f}) = \mathcal{DE}(\boldsymbol{f}|\gamma) \propto \exp\left[\gamma \|\boldsymbol{f}\|_1\right] \end{cases}$$

  - Maximum A Posteriori (MAP):

$$\widehat{\boldsymbol{f}} = \arg\max_{\boldsymbol{f}} \left\{ p(\boldsymbol{f}|\boldsymbol{g}) \right\} = \arg\min_{\boldsymbol{f}} \left\{ J(\boldsymbol{f}) \right\}$$

$$J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{\Phi}'\boldsymbol{f}\|^2 + \lambda \|\boldsymbol{f}\|_1 \quad \text{with} \quad \lambda = 2\gamma\sigma_\epsilon^2$$

Sparse decomposition: Regularization or MAP

- Regularization: $J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{\Phi}'\boldsymbol{f}\|^2 + \lambda\|\boldsymbol{f}\|_1$
- With fixed λ: Find a good optimization algorithm
- How to choose λ ?
- L-Curve, Cross Validation, adhoc $\lambda = 1$, ...
- ---
- MAP: $J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{\Phi}'\boldsymbol{f}\|^2 + \lambda\|\boldsymbol{f}\|_1$ with $\lambda = 2\gamma\sigma_\epsilon^2$
- How to estimate γ and σ_ϵ^2?
- Bayesian: Joint MAP, Expectation-Maximization, MCMC, Variational Bayesian Approximation,...
- Advantages of the Bayesian approach:
  - More probabilistic modeling for sparsity enforcing
  - Hyperparameter estimation
  - Uncertainty handling

5. Sparsity enforcing prior models

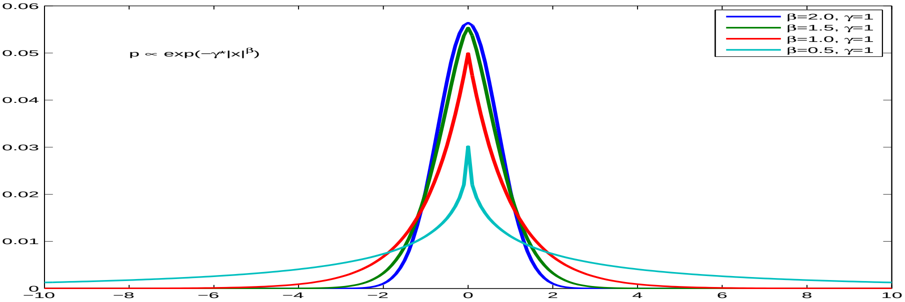
- Simple heavy tailed models:
  - Generalized Gaussian, Double Exponential
  - Student-t, Cauchy
  - Elastic net
  - Generalized hyperbolic
  - Symmetric Weibull, Symmetric Rayleigh
- Hierarchical mixture models:
  - Mixture of Gaussians
  - Bernoulli-Gaussian
  - Mixture of Gammas
  - Bernoulli-Gamma
  - Mixture of Dirichlet
  - Bernoulli-Multinomial

Simple heavy tailed models

- Generalized Gaussian, Double Exponential

$$p(\boldsymbol{f}|\gamma,\beta) = \prod_j \mathcal{GG}(f_j|\gamma,\beta) \propto \exp\left[-\gamma \sum_j |f_j|^\beta\right]$$

$\beta = 1$ Double exponential or Laplace.
$0 < \beta < 2$ are of great interest for sparsity enforcing.



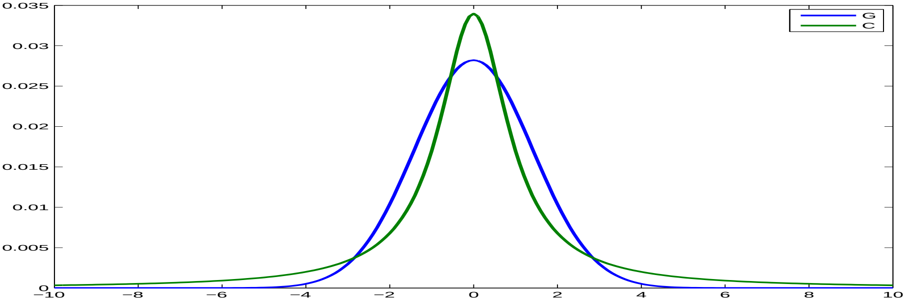
Generalized Gaussian family

Simple heavy tailed models

- Student-t and Cauchy models

$$p(\boldsymbol{f}|\nu) = \prod_j \mathcal{S}t(f_j|\nu) \propto \exp\left[-\frac{\nu+1}{2}\sum_j \log\left(1+f_j^2/\nu\right)\right]$$

Cauchy model is obtained when $\nu = 1$.

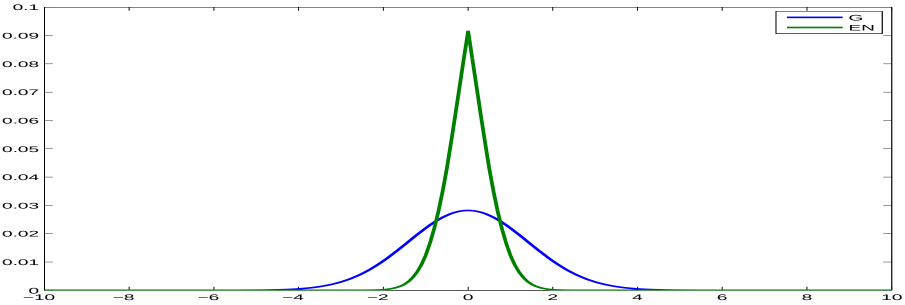


Student-t and Cauchy families

Simple heavy tailed models

- Elastic net prior model

$$p(\boldsymbol{f}|\nu) = \prod_j \mathcal{EN}(f_j|\nu) \propto \exp\left[-\sum_j (\gamma_1|f_j| + \gamma_2 f_j^2)\right]$$

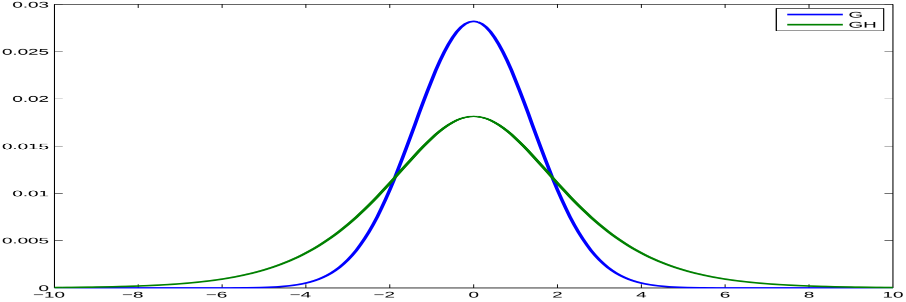


Elastic Net family

Simple heavy tailed models

- Generalized hyperbolic (GH) models

$$p(\boldsymbol{f}|\delta,\nu,\beta) = \prod_j (\delta^2 + f_j^2)^{(\nu-1/2)/2} \exp\left[\beta x\right] K_{\nu-1/2}(\alpha\sqrt{\delta^2 + f_j^2})$$



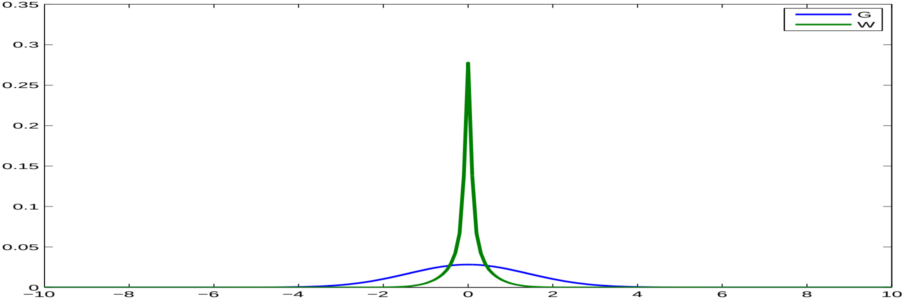
Generalized hyperbolic family

Simple heavy tailed models

- Symmetric Weibull

$$p(\boldsymbol{f}|\gamma,\beta) = \prod_j \mathcal{W}(f_j|\gamma,\beta) \propto \exp\left[-\gamma \sum_j |f_j|^\beta + (\beta-1)\log|f_j|\right]$$

$\beta = 2$ is the Symmetric Rayleigh distribution.
$\beta = 1$ is the Double exponential and
$0 < \beta < 2$ are of great interest for sparsity enforcing.



Symmetric Weibull family

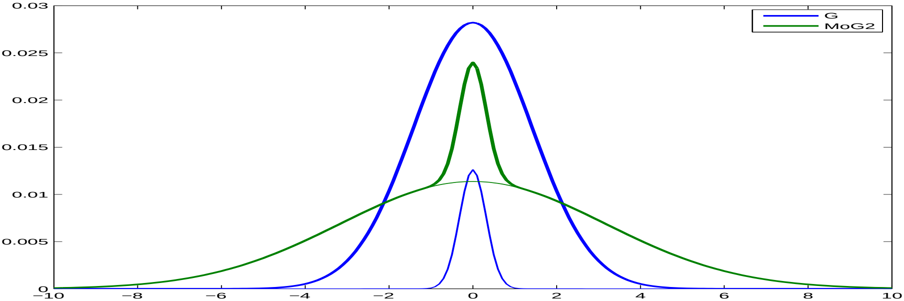
Mixture models

- Mixture of two Gaussians (MoG2) model

$$p(\boldsymbol{f}|\alpha, v_1, v_0) = \prod_j \left[\alpha \mathcal{N}(f_j|0, v_1) + (1-\alpha)\mathcal{N}(f_j|0, v_0)\right]$$

- Bernoulli-Gaussian (BG) model

$$p(\boldsymbol{f}|\alpha, v) = \prod_j p(f_j) = \prod_j \left[\alpha \mathcal{N}(f_j|0, v) + (1-\alpha)\delta(f_j)\right]$$



Mixture of 2 Gaussians families

- Mixture of Gammas

$$p(\boldsymbol{f}|\lambda, v_1, v_0) = \prod_j \left[\lambda \mathcal{G}(f_j|\alpha_1, \beta_1) + (1-\lambda)\mathcal{G}(f_j|\alpha_2, \beta_2)\right]$$

- Bernoulli-Gamma model

$$p(\boldsymbol{f}|\lambda, \alpha, \beta) = \prod_j \left[\lambda \mathcal{G}(f_j|\alpha, \beta) + (1-\lambda)\delta(f_j)\right]$$

- Mixture of Dirichlets model

$$p(\boldsymbol{f}|\lambda, \boldsymbol{H}_1, \boldsymbol{\alpha}_1, \boldsymbol{H}_2, \boldsymbol{\alpha}_2) = \prod_j \left[\lambda \mathcal{D}(f_j|\boldsymbol{H}_1, \boldsymbol{\alpha}_1) + (1-\lambda)\mathcal{D}(f_j|\boldsymbol{H}_2, \boldsymbol{\alpha}_2)\right]$$

$$\mathcal{D}(f_j|\boldsymbol{H}, \boldsymbol{\alpha}) = \prod_{k=1}^{K} \frac{\Gamma(\alpha)}{\Gamma(\alpha_0)\Gamma(\alpha_K)} a_k^{\alpha_k - 1}, \quad \alpha_k \geq 0, \quad a_k \geq 0$$

where $\boldsymbol{H} = \{a_1, \cdots, a_K\}$ and $\boldsymbol{\alpha} = \{\alpha_1, \cdots, \alpha_K\}$
with $\sum_k \alpha_k = \alpha$ and $\sum_k a_k = 1$.

- Bernoulli-Multinomial (BMultinomial) model

$$p(\boldsymbol{f}|\lambda, \boldsymbol{H}, \boldsymbol{\alpha}) = \prod_j \left[\lambda \delta(f_j) + (1-\lambda)\mathcal{M}ult(f_j|\boldsymbol{H}, \boldsymbol{\alpha})\right]$$

Hierarchical models and hidden variables

- All the mixture models and some of simple models can be modeled via hidden variables $\boldsymbol{z}$.

$$p(f) = \sum_{k=1}^{K} \alpha_k p_k(f) \longrightarrow \left\{ \begin{array}{l} p(f|z=k) = p_k(f), \\ P(z=k) = \alpha_k, \quad \sum_k \alpha_k = 1 \end{array} \right.$$

- Example 1: MoG model: $p_k(f) = \mathcal{N}(f|m_k, v_k)$
  2 Gaussians: $p_0 = \mathcal{N}(0, v_0), p_1 = \mathcal{N}(0, v_1), \alpha_0 = \lambda, \alpha_1 = 1 - \lambda$

$$p(f_j|\lambda, v_1, v_0) = \lambda \mathcal{N}(f_j|0, v_1) + (1-\lambda)\mathcal{N}(f_j|0, v_0)$$

$$\left\{ \begin{array}{l} p(f_j|z_j = 0, v_0) = \mathcal{N}(f_j|0, v_0), \\ p(f_j|z_j = 1, v_1) = \mathcal{N}(f_j|0, v_1), \end{array} \right. \text{ and } \left\{ \begin{array}{l} P(z_j = 0) = \lambda, \\ P(z_j = 1) = 1 - \lambda \end{array} \right.$$

$$\left\{ \begin{array}{l} p(\boldsymbol{f}|\boldsymbol{z}) = \prod_j p(f_j|z_j) = \prod_j \mathcal{N}\left(f_j|0, v_{z_j}\right) \propto \exp\left[-\frac{1}{2}\sum_j \frac{f_j^2}{v_{z_j}}\right] \\ p(\boldsymbol{z}) = \lambda^{n_1}(1-\lambda)^{n_0}, \quad n_1 = \sum_j \delta(z_j - 1), \quad n_0 = \sum_j \delta(z_j) \end{array} \right.$$

Hierarchical models and hidden variables

- Example 2: Student-t model

$$\mathcal{S}t(f|\nu) \propto \exp\left[-\frac{\nu+1}{2}\log\left(1+f^2/\nu\right)\right]$$

- Infinite mixture

$$\mathcal{S}t(f|\nu) \propto= \int_0^\infty \mathcal{N}(f|,0,1/z)\,\mathcal{G}(z|\alpha,\beta)\;\mathrm{d}z, \quad \text{with } \alpha=\beta=\nu/2$$

$$p(f|z)=\mathcal{N}(f|0,1/z), \quad p(z)=\mathcal{G}(z|\alpha,\beta)$$

$$\begin{cases} p(\boldsymbol{f}|\boldsymbol{z}) & = \prod_j p(f_j|z_j) = \prod_j \mathcal{N}(f_j|0,1/z_j) \propto \exp\left[-\frac{1}{2}\sum_j z_j f_j^2\right] \\ p(\boldsymbol{z}|\alpha,\beta) & = \prod_j \mathcal{G}(z_j|\alpha,\beta) \propto \prod_j {z_j}^{(\alpha-1)} \exp\left[-\beta z_j\right] \\ & \propto \exp\left[\sum_j (\alpha-1)\ln z_j - \beta z_j\right] \\ p(\boldsymbol{f},\boldsymbol{z}|\alpha,\beta) & \propto \exp\left[-\frac{1}{2}\sum_j z_j f_j^2 + (\alpha-1)\ln z_j - \beta z_j\right] \end{cases}$$

Inverse Problems

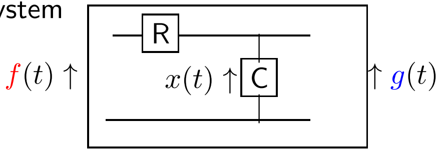
Three main examples

- Example 1:
  Instrumentation
  - $f(t)$ input of the instrument
  - $g(t)$ output of the instrument
- Example 2: Seeing outside of a body: Making an image using a camera, a microscope or a telescope
  - $f(x, y)$ real scene
  - $g(x, y)$ observed image
- Example 3: Seeing inside of a body: Computed Tomography usng X rays, US, Microwave, etc.
  - $f(x, y)$ a section of a real 3D body $f(x, y, z)$
  - $g_\phi(r)$ a line of observed radiographe $g_\phi(r, z)$

- Example 1: Deconvolution
- Example 2: Image restoration
- Example 3: Image reconstruction

Instrumentation

A simple electric system



$$f(t) = R\,i(t) + v_c(t) = RC\,\frac{\partial x(t)}{\partial t} + x(t), \quad RC = 1$$

- Differential Equation Modelling

$$\frac{\partial x(t)}{\partial t} + x(t) = f(t), \qquad x(t) = g(t)$$

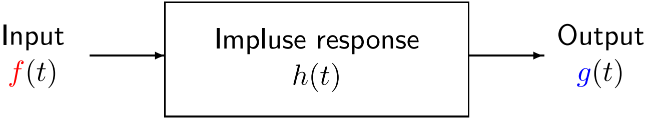
- State Space Modelling

$$\begin{cases} \frac{\partial x(t)}{\partial t} & = -x(t) + f(t) \\ g(t) & = x(t) \end{cases}$$

- Input-Output Modelling

$$\begin{cases} \frac{\partial x(t)}{\partial t} = -x(t) + f(t) \\ g(t) = x(t) \end{cases} \rightarrow \begin{cases} pX(p) = -X(p) + F(p) \rightarrow X(p) = \frac{1}{p+1} F(p) \\ g(t) = x(t) = h(t) * f(t), \quad h(t) = \exp[-t] \end{cases}$$

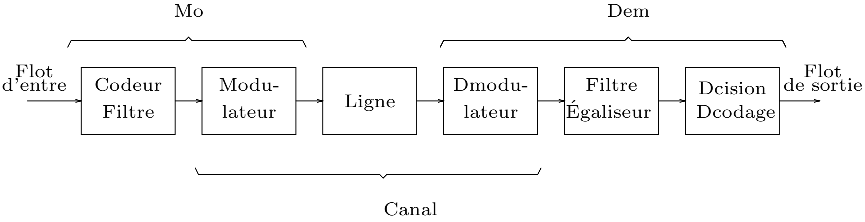
Instrumentation



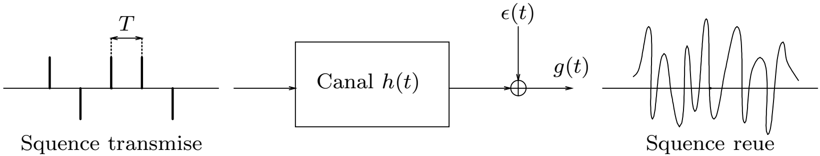
- Ideal Instrument $\quad g(t) = f(t) \quad$ does not exist.
- A linear and time invariant instrument is characterized by its impulse response $\quad h(t)$.
- Ideal Instrument $\quad h(t) = \delta(t) \quad$ does not exist.
- Forward problem: $\quad f(t),\, h(t) \longrightarrow g(t) = h(t) * f(t)$
- Two linked problems in instrumentation:
  - Inversion: $\quad g(t),\, h(t) \longrightarrow f(t)$
  - Identification: $\quad g(t),\, f(t) \longrightarrow h(t)$

Telecommunication: transmission channel compensation

- Data transmission System

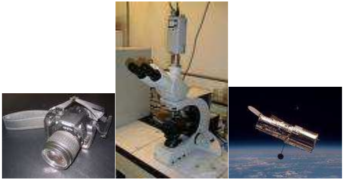


- Channel Model: convolution + noise



Seeing outside of a body: Making an image with a camera, a microscope or a telescope

- $f(x, y)$ real scene
- $g(x, y)$ observed image



- Forward model: Convolution

$$g(x, y) = \iint f(x', y') \, h(x - x', y - y') \, \mathrm{d}x' \, \mathrm{d}y' + \epsilon(x, y)$$

  $h(x, y)$: Point Spread Function (PSF) of the imaging system

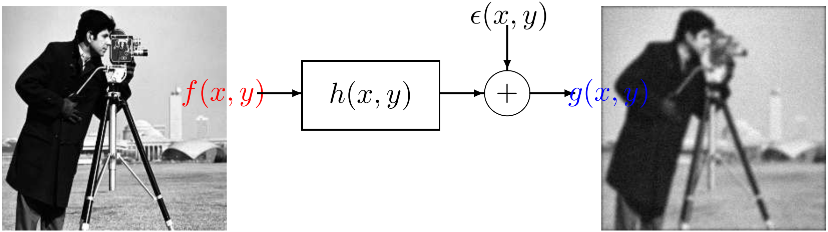
- Inverse problem: Image restoration

  Given the forward model $\mathcal{H}$ (PSF $h(x, y)$))
  and a set of data $g(x_i, y_i), i = 1, \cdots, M$
  find $f(x, y)$

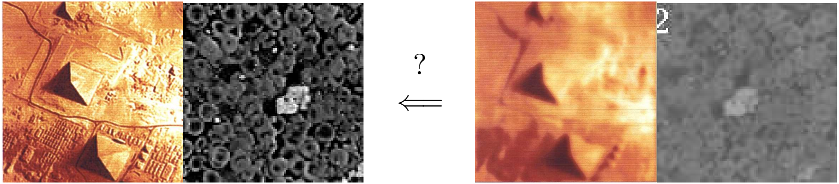
Making an image with an unfocused camera

Forward model: 2D Convolution

$$g(x,y) = \iint f(x',y')\, h(x-x', y-y')\, \mathrm{d}x'\, \mathrm{d}y' + \epsilon(x,y)$$

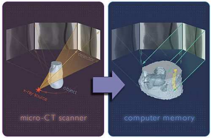


Inversion: Image Deconvolution or Restoration



Seeing inside of a body: Computed Tomography

- $f(x, y)$ a section of a real 3D body $f(x, y, z)$
- $g_\phi(r)$ a line of observed radiographe $g_\phi(r, z)$



- Forward model:
  Line integrals or Radon Transform

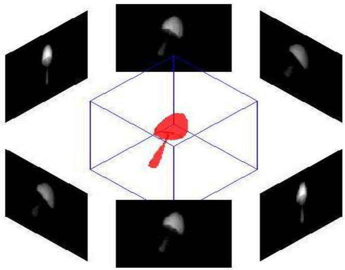
$$
\begin{aligned}
g_\phi(r) &= \int_{L_{r,\phi}} f(x, y) \, \mathrm{d}l + \epsilon_\phi(r) \\
&= \iint f(x, y) \, \delta(r - x \cos\phi - y \sin\phi) \, \mathrm{d}x \, \mathrm{d}y + \epsilon_\phi(r)
\end{aligned}
$$

- Inverse problem: Image reconstruction

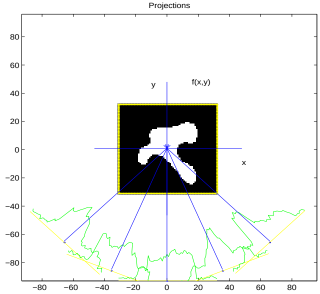
  Given the forward model $\mathcal{H}$ (Radon Transform) and
  a set of data $g_{\phi_i}(r), i = 1, \cdots, M$
  find $f(x, y)$

2D and 3D Computed Tomography

3D



2D

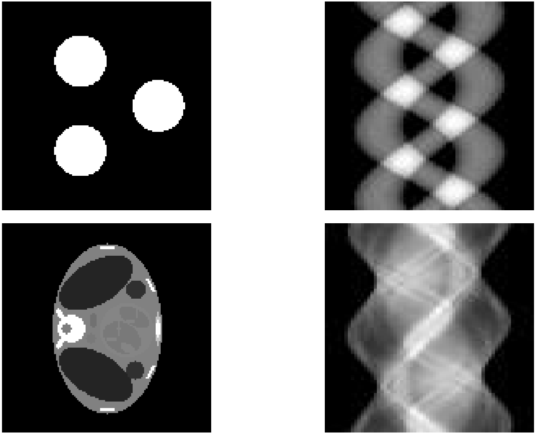


$$g_\phi(r_1, r_2) = \int_{\mathcal{L}_{r_1,r_2,\phi}} f(x, y, z) \,\mathrm{d}l \qquad g_\phi(r) = \int_{\mathcal{L}_{r,\phi}} f(x, y) \,\mathrm{d}l$$

Forward probelm: $f(x, y)$ or $f(x, y, z) \longrightarrow g_\phi(r)$ or $g_\phi(r_1, r_2)$

Inverse problem: $g_\phi(r)$ or $g_\phi(r_1, r_2) \longrightarrow f(x, y)$ or $f(x, y, z)$

Computed Tomography: Radon Transform



Forward:	$f(x, y)$	$\longrightarrow$	$g(r, \phi)$
Inverse:	$f(x, y)$	$\longleftarrow$	$g(r, \phi)$

Linear Inverse Problems

$$g(\boldsymbol{s}_i) = \int h(\boldsymbol{s}_i, \boldsymbol{r})\, f(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r} + \epsilon(\boldsymbol{s}_i), \quad i = 1, \cdots, M$$

- $f(\boldsymbol{r})$ is assumed to be well approximated by

$$f(\boldsymbol{r}) \simeq \sum_{j=1}^{N} f_j\, \phi_j(\boldsymbol{r})$$

with $\{\phi_j(\boldsymbol{r})\}$ a basis or any other set of known functions

$$g(\boldsymbol{s}_i) = g_i \simeq \sum_{j=1}^{N} f_j \int h(\boldsymbol{s}_i, \boldsymbol{r})\, \phi_j(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r}, \quad i = 1, \cdots, M$$

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon} \;\text{ with }\; H_{ij} = \int h(\boldsymbol{s}_i, \boldsymbol{r})\, \phi_j(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r}$$

- $\boldsymbol{H}$ is huge dimensional
- Regularization : $\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{J(\boldsymbol{f})\}$ with

$$J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \lambda \|\boldsymbol{f}\|_1$$

Linear Inverse problems with Hierarchical prior models

Linear inverse problems:

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$$

Likelihood:

$$p(\boldsymbol{g}|\boldsymbol{f}) = \mathcal{N}(\boldsymbol{g}|\boldsymbol{H}\boldsymbol{f}, \sigma_\epsilon^2 \boldsymbol{I})$$

- Simple priors

$$p(\boldsymbol{f}|\boldsymbol{g}, \boldsymbol{\theta}) \propto p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta}_1)\, p(\boldsymbol{f}|\boldsymbol{\theta}_2)$$

- Hierarchical priors

$$p(\boldsymbol{f}, \boldsymbol{z}|\boldsymbol{g}, \boldsymbol{\theta}) \propto p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta}_1)\, p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\theta}_2)\, p(\boldsymbol{z}|\boldsymbol{\theta}_3)$$

$$\left\{ \begin{array}{ll} p(\boldsymbol{f}|\boldsymbol{z}) & = \mathcal{N}(\boldsymbol{f}|\boldsymbol{\Phi}'\boldsymbol{z}, \sigma_f^2 \boldsymbol{I}) \\ p(\boldsymbol{z}) & = \prod_j p(z_j) \end{array} \right.$$

Linear Inverse problems with Hierarchical prior models

Linear inverse problems:

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$$

Likelihood:

$$p(\boldsymbol{g}|\boldsymbol{f}) = \mathcal{N}(\boldsymbol{g}|\boldsymbol{H}\boldsymbol{f}, \sigma_\epsilon^2 \boldsymbol{I})$$

- With Hyperparameters $\boldsymbol{\theta}$ we have:
  - Simple priors

$$p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta}_1)\, p(\boldsymbol{f}|\boldsymbol{\theta}_2)\, p(\boldsymbol{\theta})$$

  - Hierarchical priors

$$p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta}_1)\, p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\theta}_2)\, p(\boldsymbol{z}|\boldsymbol{\theta}_3)\, p(\boldsymbol{\theta})$$

Bayesian Computation and Algorithms

- When the expression of $p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g})$ or of $p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{g})$ is obtained, we have following options:
- Joint MAP: (needs optimization algorithms)

$$(\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{\theta}}) = \arg\max_{(\boldsymbol{f},\boldsymbol{\theta})} \{p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g})\}$$

- MCMC: Needs the expressions of the conditionals $p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{g})$, $p(\boldsymbol{z}|\boldsymbol{f}, \boldsymbol{\theta}, \boldsymbol{g})$, and $p(\boldsymbol{\theta}|\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{g})$
- Variational Bayesian Approximation (VBA): Approximate $p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{g})$ by a separable one

$$q(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{g}) = q_1(\boldsymbol{f})\, q_2(\boldsymbol{z})\, q_3(\boldsymbol{\theta})$$

  and do any computations with these separable ones.

Joint MAP

$$p(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{\theta}_1)\; p(\boldsymbol{f}|\boldsymbol{\theta}_2)\; p(\boldsymbol{\theta})$$

- Objective:

$$(\widehat{\boldsymbol{f}},\widehat{\boldsymbol{\theta}}) = \arg\max_{(\boldsymbol{f},\boldsymbol{\theta})} \{p(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g})\}$$

- Alternate optimization:

$$\left\{\begin{array}{l} \widehat{\boldsymbol{f}} = \arg\max_{\boldsymbol{f}} \left\{p(\boldsymbol{f},\widehat{\boldsymbol{\theta}}|\boldsymbol{g})\right\} \\ \widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \left\{p(\widehat{\boldsymbol{f}},\boldsymbol{\theta}|\boldsymbol{g})\right\} \end{array}\right.$$

$$\begin{array}{ccccc} \boldsymbol{\theta}^{(0)} \longrightarrow & \widehat{\boldsymbol{\theta}} \longrightarrow & \boxed{\widehat{\boldsymbol{f}} = \arg\max_{\boldsymbol{f}} \left\{p(\boldsymbol{f}|\widehat{\boldsymbol{\theta}},\boldsymbol{g})\right\}} & \longrightarrow & \widehat{\boldsymbol{f}} \\ & \uparrow & & & \downarrow \\ & \widehat{\boldsymbol{\theta}} \longleftarrow & \boxed{\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \left\{p(\boldsymbol{\theta}|\widehat{\boldsymbol{f}},\boldsymbol{g})\right\}} & \longleftarrow & \widehat{\boldsymbol{f}} \end{array}$$

- Uncertainties are not propagated.

MCMC based algorithm

$$p(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}|\boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta})\, p(\boldsymbol{f}|\boldsymbol{z},\boldsymbol{\theta})\, p(\boldsymbol{z})\, p(\boldsymbol{\theta})$$

General scheme (Gibbs Sampling):

- Generate samples from the conditionals:

$$\widehat{\boldsymbol{f}} \sim p(\boldsymbol{f}|\widehat{\boldsymbol{z}},\widehat{\boldsymbol{\theta}},\boldsymbol{g}) \longrightarrow \widehat{\boldsymbol{z}} \sim p(\boldsymbol{z}|\widehat{\boldsymbol{f}},\widehat{\boldsymbol{\theta}},\boldsymbol{g}) \longrightarrow \widehat{\boldsymbol{\theta}} \sim (\boldsymbol{\theta}|\widehat{\boldsymbol{f}},\widehat{\boldsymbol{z}},\boldsymbol{g})$$

- Waite for convergency
- Compute empirical statistics (means, modes, variances) from the samples

$$\mathsf{E}\left\{\boldsymbol{f}\right\} \approx \frac{1}{N}\sum_{n} \boldsymbol{f}^{(n)}$$

Variational Bayesian Approximation

- Approximate $p(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g})$ by $q(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g}) = q_1(\boldsymbol{f}|\boldsymbol{g})\, q_2(\boldsymbol{\theta}|\boldsymbol{g})$ and then continue computations.
- Criterion $\mathrm{KL}(q(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g}) : p(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g}))$

$$\mathrm{KL}(q:p) = \iint q \ln \frac{q}{p} = \iint q_1 q_2 \ln \frac{q_1 q_2}{p}$$

- Iterative algorithm $q_1 \longrightarrow q_2 \longrightarrow q_1 \longrightarrow q_2, \cdots$

$$\left\{ \begin{array}{ll} q_1(\boldsymbol{f}) & \propto \exp\left[\langle \ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta};\mathcal{M})\rangle_{q_2(\boldsymbol{\theta})}\right] \\ q_2(\boldsymbol{\theta}) & \propto \exp\left[\langle \ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta};\mathcal{M})\rangle_{q_1(\boldsymbol{f})}\right] \end{array} \right.$$

$$\widehat{q}_2^{(0)} \longrightarrow \widehat{q}_2 \longrightarrow \boxed{q_1(\boldsymbol{f}) \propto \exp\left[\langle \ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta};\mathcal{M})\rangle_{q_2}\right]} \longrightarrow \widehat{q}_1(\boldsymbol{f}) \longrightarrow \widehat{\boldsymbol{f}}$$

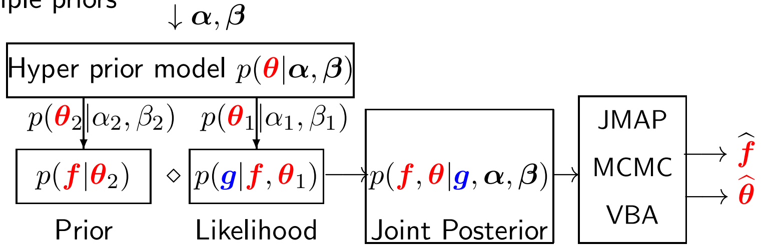
$$\uparrow \qquad\qquad\qquad\qquad \downarrow$$

$$\widehat{\boldsymbol{\theta}} \leftarrow \widehat{q}_2(\boldsymbol{\theta}) \longleftarrow \boxed{q_2(\boldsymbol{\theta}) \propto \exp\left[\langle \ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta};\mathcal{M})\rangle_{q_1}\right]} \longleftarrow \widehat{q}_1$$

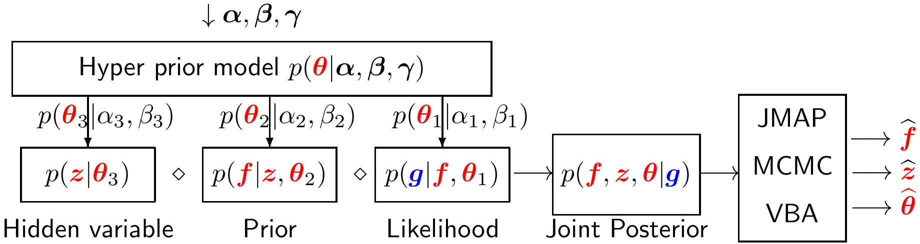
- Uncertainties are propagated (Message Passing methods)

Summary of Bayesian approach

- Simple priors



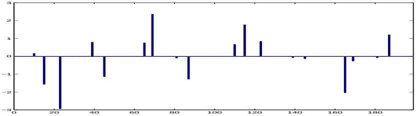
- Hierarchical priors



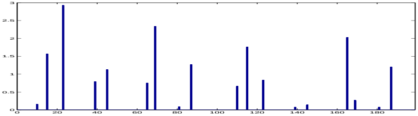
Advantages of the Bayesian Approach

- More possibilities to model sparsity
- More tools to handle hyperparameters
- More tools to account for uncertainties
- More possibilities to understand and to control many ad hoc deterministic algorithms
- Hierarchical models give still more modeling possibilities
  - Bernouilli-Gaussian: strict sparsity
  - Bernouilli-Gamma: strict sparsity + positivity
  - Bernouilli-Multinomial: strict sparsity + discrete values (finite states)
  - Independent Mixture models: sparsity enforcing
  - Mixture of multivariate models: group sparsity enforcing
  - Gauss-Markov-Potts models: sparsity in transform domains

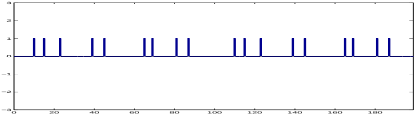
Examples of Hierarchical models



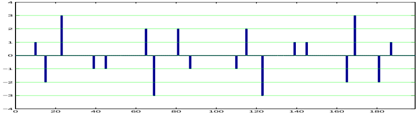
Bernouilli-Gaussian



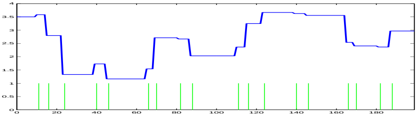
Bernouilli-Gamma



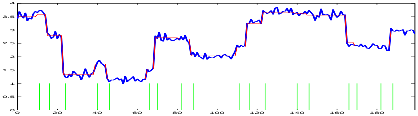
Bernouilli-Binomial



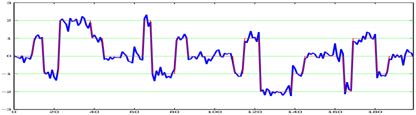
Bernouilli-Multinomial



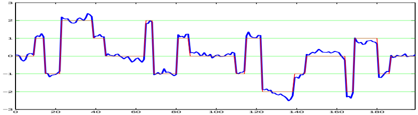
Piecewise constant



Piecewise Gaussian

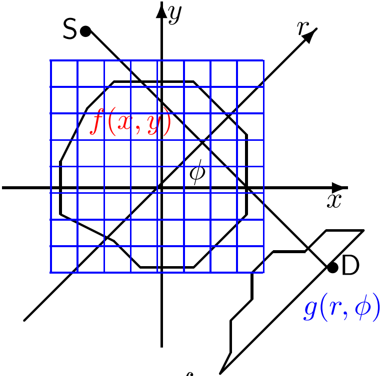


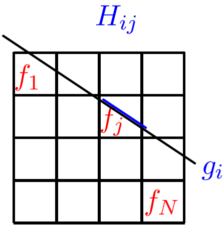
Gauss-Markov-Potts 1



Gauss-Markov-Potts 2

Computed Tomography: Discretization





$$f(x,y) = \sum_j f_j \, b_j(x,y)$$

$$b_j(x,y) = \begin{cases} 1 & \text{if } (x,y) \in \text{ pixel } j \\ 0 & \text{else} \end{cases}$$

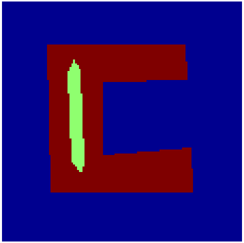
$$g(r,\phi) = \int_L f(x,y) \, \mathrm{d}l$$

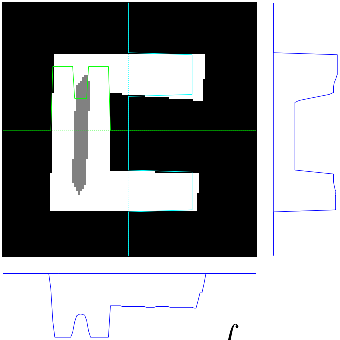
$$g_i = \sum_{j=1}^{N} H_{ij} \, f_j + \epsilon_i$$

$$\boldsymbol{g} = \boldsymbol{H} \boldsymbol{f} + \boldsymbol{\epsilon}$$

Application of CT in NDT

Reconstruction from only 2 projections



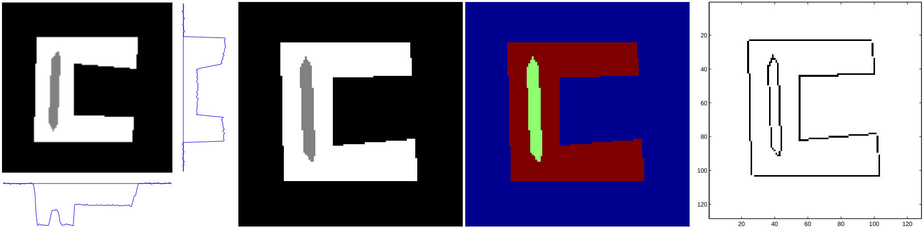


$$g_1(x) = \int f(x,y) \, \mathrm{d}y, \qquad g_2(y) = \int f(x,y) \, \mathrm{d}x$$

- Given the marginals $g_1(x)$ and $g_2(y)$ find the joint distribution $f(x,y)$.
- Infinite number of solutions : $f(x,y) = g_1(x)\, g_2(y)\, \Omega(x,y)$
  $\Omega(x,y)$ is a Copula:

$$\int \Omega(x,y) \, \mathrm{d}x = 1 \quad \text{and} \quad \int \Omega(x,y) \, \mathrm{d}y = 1$$

Application in CT



$\boldsymbol{g}\vert\boldsymbol{f}$	$\boldsymbol{f}\vert\boldsymbol{z}$	$\boldsymbol{z}$	$\boldsymbol{c}$
$\boldsymbol{g}=\boldsymbol{H}\boldsymbol{f}+\boldsymbol{\epsilon}$	iid Gaussian	iid	$q(\boldsymbol{r})\in\{0,1\}$
$\boldsymbol{g}\vert\boldsymbol{f}\sim\mathcal{N}(\boldsymbol{H}\boldsymbol{f},\sigma_\epsilon^2\boldsymbol{I})$	or	or	$1-\delta(z(\boldsymbol{r})-z(\boldsymbol{r}'))$
Gaussian	Gauss-Markov	Potts	binary

$$p(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}|\boldsymbol{g})\propto p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{\theta}_1)\,p(\boldsymbol{f}|\boldsymbol{z},\boldsymbol{\theta}_2)\,p(\boldsymbol{z}|\boldsymbol{\theta}_3)\,p(\boldsymbol{\theta})$$

Proposed algorithms

$$p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta}_1)\, p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\theta}_2)\, p(\boldsymbol{z}|\boldsymbol{\theta}_3)\, p(\boldsymbol{\theta})$$

- MCMC based general scheme:

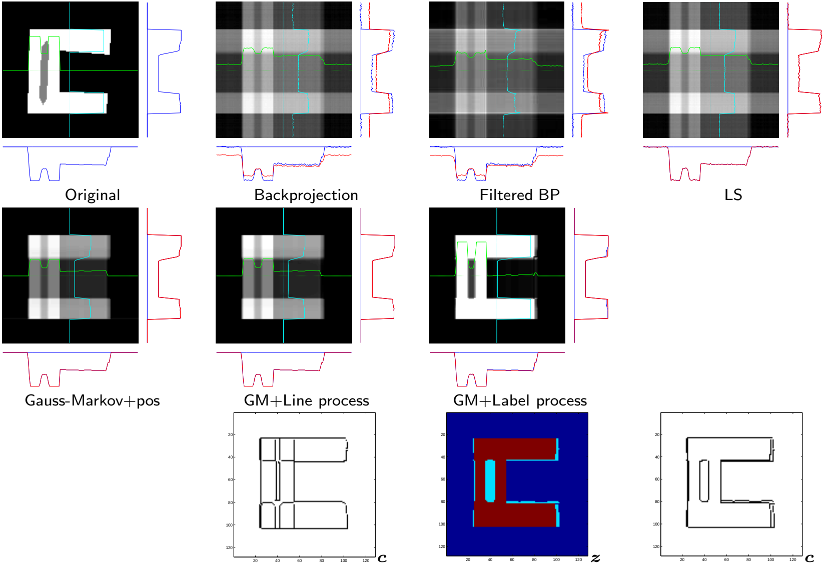
$$\widehat{\boldsymbol{f}} \sim p(\boldsymbol{f}|\widehat{\boldsymbol{z}}, \widehat{\boldsymbol{\theta}}, \boldsymbol{g}) \longrightarrow \widehat{\boldsymbol{z}} \sim p(\boldsymbol{z}|\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{\theta}}, \boldsymbol{g}) \longrightarrow \widehat{\boldsymbol{\theta}} \sim (\boldsymbol{\theta}|\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{z}}, \boldsymbol{g})$$

Iterative algorithme:

- Estimate $\boldsymbol{f}$ using $p(\boldsymbol{f}|\widehat{\boldsymbol{z}}, \widehat{\boldsymbol{\theta}}, \boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta})\, p(\boldsymbol{f}|\widehat{\boldsymbol{z}}, \widehat{\boldsymbol{\theta}})$
  Needs optimisation of a quadratic criterion.
- Estimate $\boldsymbol{z}$ using $p(\boldsymbol{z}|\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{\theta}}, \boldsymbol{g}) \propto p(\boldsymbol{g}|\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{z}}, \widehat{\boldsymbol{\theta}})\, p(\boldsymbol{z})$
  Needs sampling of a Potts Markov field.
- Estimate $\boldsymbol{\theta}$ using
  $p(\boldsymbol{\theta}|\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{z}}, \boldsymbol{g}) \propto p(\boldsymbol{g}|\widehat{\boldsymbol{f}}, \sigma_\epsilon^2 \boldsymbol{I})\, p(\widehat{\boldsymbol{f}}|\widehat{\boldsymbol{z}}, (m_k, v_k))\, p(\boldsymbol{\theta})$
  Conjugate priors $\longrightarrow$ analytical expressions.

- Variational Bayesian Approximation
  - Approximate $p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{g})$ by $q_1(\boldsymbol{f})\, q_2(\boldsymbol{z})\, q_3(\boldsymbol{\theta})$

Results

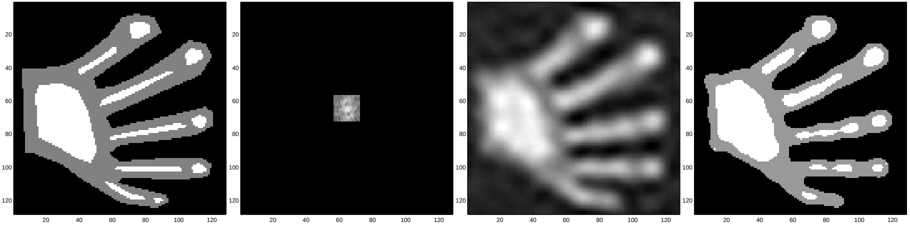

Original
Backprojection
Filtered BP
LS
Gauss-Markov+pos
GM+Line process
GM+Label process
c
z
c

Application in Microwave imaging

$$g(\boldsymbol{\omega}) = \int f(\boldsymbol{r}) \exp\left[-j(\boldsymbol{\omega}.\boldsymbol{r})\right] \, \mathrm{d}\boldsymbol{r} + \epsilon(\boldsymbol{\omega})$$

$$g(u,v) = \iint f(x,y) \exp\left[-j(ux+vy)\right] \, \mathrm{d}x \, \mathrm{d}y + \epsilon(u,v)$$

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$$



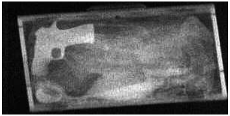
$f(x,y)$ $g(u,v)$ $\widehat{\boldsymbol{f}}$ IFT $\widehat{\boldsymbol{f}}$ Proposed method

Images fusion and joint segmentation

(with O. Féron)

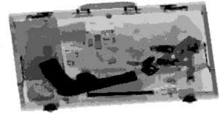
$$\begin{cases} g_i(\boldsymbol{r}) = f_i(\boldsymbol{r}) + \epsilon_i(\boldsymbol{r}) \\ p(f_i(\boldsymbol{r}) | z(\boldsymbol{r}) = k) = \mathcal{N}(m_{ik}, \sigma^2_{i\,k}) \\ p(\underline{\boldsymbol{f}} | \boldsymbol{z}) = \prod_i p(\boldsymbol{f}_i | \boldsymbol{z}) \end{cases}$$

$\boldsymbol{g}_1$

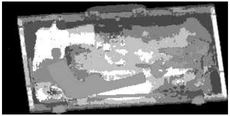


$\boldsymbol{g}_2$

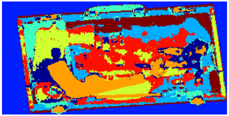
$\longrightarrow$



$\widehat{\boldsymbol{f}}_1$



$\widehat{\boldsymbol{f}}_2$

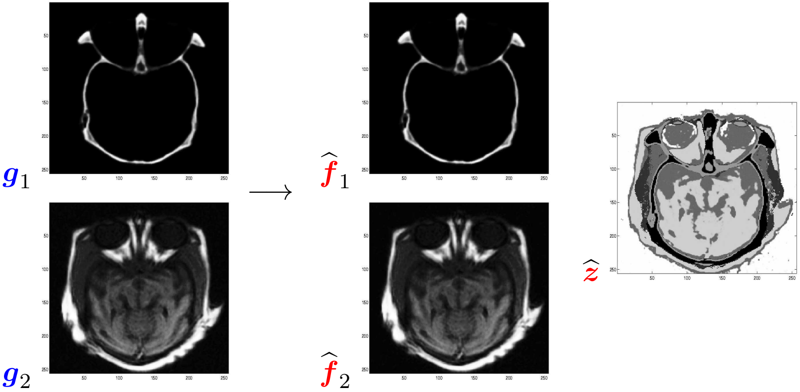


$\widehat{\boldsymbol{z}}$

Data fusion in medical imaging

(with O. Féron)

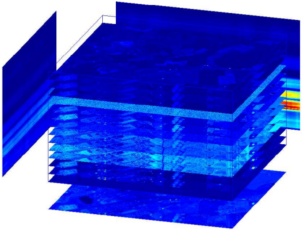
$$
\begin{cases}
g_i(\boldsymbol{r}) = f_i(\boldsymbol{r}) + \epsilon_i(\boldsymbol{r}) \\
p(f_i(\boldsymbol{r}) | z(\boldsymbol{r}) = k) = \mathcal{N}(m_{i\,k}, \sigma_{i\,k}^2) \\
p(\underline{\boldsymbol{f}} | \boldsymbol{z}) = \prod_i p(\boldsymbol{f}_i | \boldsymbol{z})
\end{cases}
$$

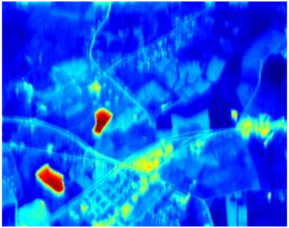


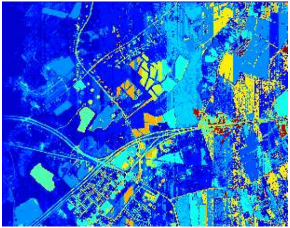
Joint segmentation of hyper-spectral images

(with N. Bali & A. Mohammadpour)

$$\begin{cases} g_i(\boldsymbol{r}) = f_i(\boldsymbol{r}) + \epsilon_i(\boldsymbol{r}) \\ p(f_i(\boldsymbol{r})|z(\boldsymbol{r}) = k) = \mathcal{N}(m_{ik}, \sigma^2_{i\,k}), \quad k = 1, \cdots, K \\ p(\underline{\boldsymbol{f}}|\boldsymbol{z}) = \prod_i p(\boldsymbol{f}_i|\boldsymbol{z}) \\ m_{ik} \text{ follow a Markovian model along the index } i \end{cases}$$







Segmentation of a video sequence of images

(with P. Brault)

$$\begin{cases} g_i(\boldsymbol{r}) = f_i(\boldsymbol{r}) + \epsilon_i(\boldsymbol{r}) \\ p(f_i(\boldsymbol{r})|z_i(\boldsymbol{r}) = k) = \mathcal{N}(m_{i\,k}, \sigma_{i\,k}^2), \quad k = 1, \cdots, K \\ p(\underline{\boldsymbol{f}}|\boldsymbol{z}) = \prod_i p(\boldsymbol{f}_i|\boldsymbol{z}_i) \\ z_i(\boldsymbol{r}) \;\; \text{follow a Markovian model along the index} \;\; i \end{cases}$$

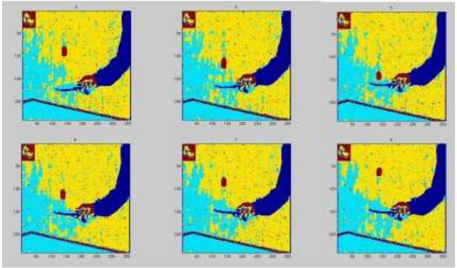
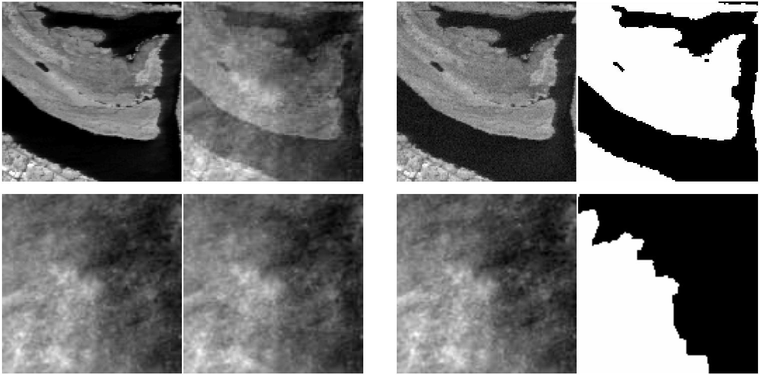


Image separation in Sattelite imaging

(with H. Snoussi & M. Ichir)

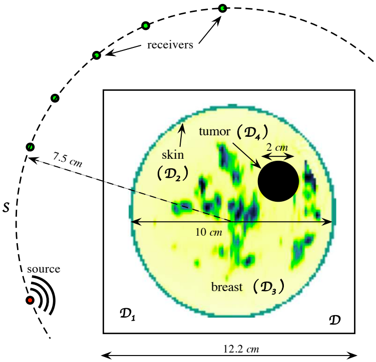
$$\begin{cases} g_i(\boldsymbol{r}) = \sum_{j=1}^{N} A_{ij} f_j(\boldsymbol{r}) + \epsilon_i(\boldsymbol{r}) \\ p(f_j(\boldsymbol{r}) | z_j(\boldsymbol{r}) = k) = \mathcal{N}(m_{j\,k}, \sigma^2_{j\,k}) \\ p(A_{ij}) = \mathcal{N}(A_{0ij}, \sigma^2_{0ij}) \end{cases}$$



$\boldsymbol{f}$ $\boldsymbol{g}$ $\widehat{\boldsymbol{f}}$ $\widehat{\boldsymbol{z}}$

Microwave imaging for breast cancer detection

(with L. Gharsally and B. Duchêne)

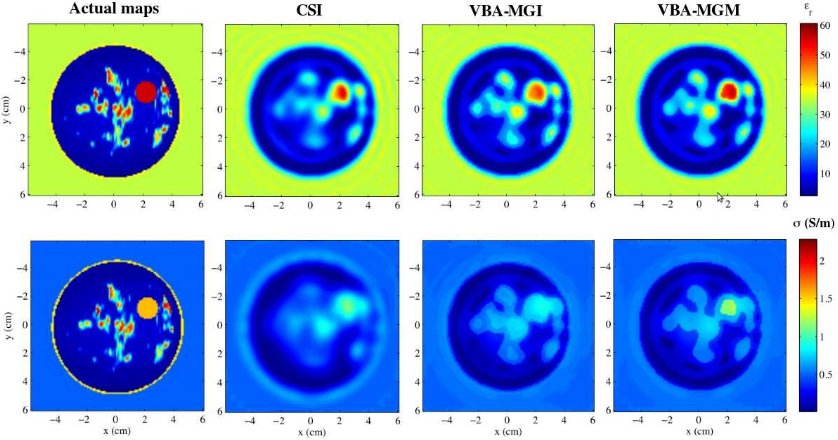


- Breast model built up from MRI scan [Zastrow et al. 2008]
- 64 sources
- 64 receivers
- 6 frequencies in the band $0.5 - 3$ GHz

$\mathcal{D} = N_x \times N_y$	D1 $(\epsilon_1, \sigma_1(S/m))$	D2 $(\epsilon_2, \sigma_2(S/m))$	D3 $(\epsilon_3, \sigma_3(S/m))$	D4 $(\epsilon_3, \sigma_3(S/m))$
120×120	$(10, 0.5)$	$(6.12, 0.11)$	$([2.46, 60.6], [0.01, 2.28])$	$(55.3, 1.57)$

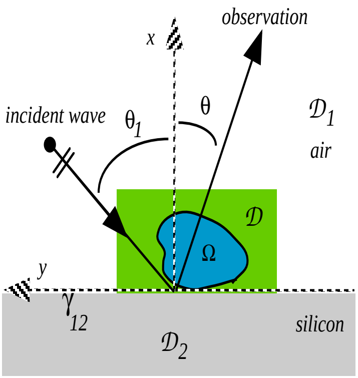
Microwave imaging for breast cancer detection

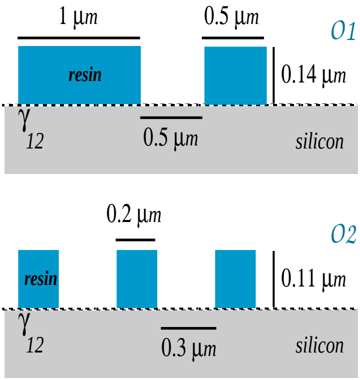
CSI: Contrast Source Inversion, VBA: Variational Bayesian Approach, MGI: Independent Gaussian mixture, MGM: Gauss-Markov mixture



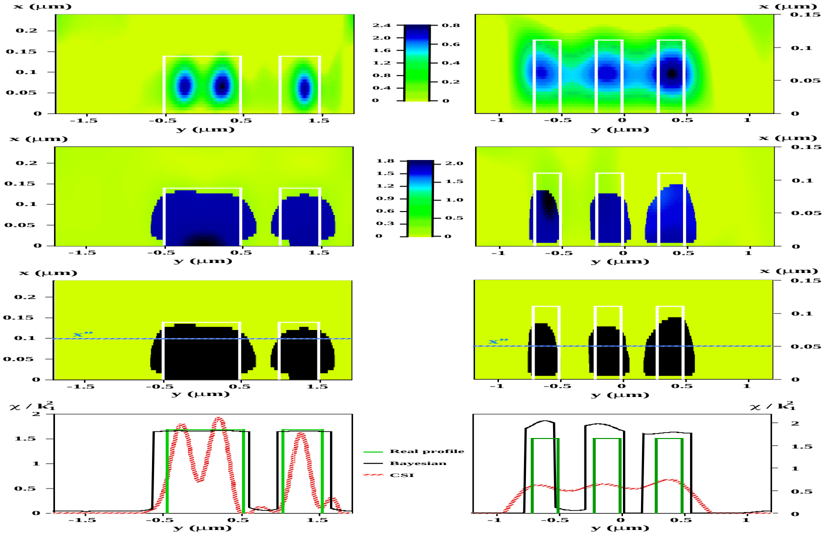
Optical Diffraction Tomographic imaging

(with H. Ayasso and B. Duchêne)





Optical Diffraction Tomographic imaging
(with H. Ayasso and B. Duchêne)



Conclusions

- Sparsity: a great property to use in signal and image processing
- Origine: Sampling theory and reconstruction, modeling and representation Compressed Sensing, Approximation theory
- Deterministic Algorithms: Optimization of a two termes criterion, penalty term, regularization term
- Probabilistic: Bayesian approach
- Sprasity enforcing priors: Simple heavy tailed and Hierarchical with hidden variables.
- Gauss-Markov-Potts models for images incorporating hidden regions and contours
- Main Bayesian computation tools: JMAP, MCMC and VBA
- Application in different imaging system (X ray CT, Microwaves, PET, ultrasound and microwave imaging)

Current Projects:

- Efficient implementation in 2D and 3D cases
- Comparison between MCMC and VBA methods

Thanks to:

Present PhD students:

- L. Gharsali (Microwave imaging for Cancer detection)
- M. Dumitru (Multivariate time series analysis for biological signals)
- S. AlAli (Diffraction imaging for geophysical applications)

Freshly Graduated PhD students:

- C. Cai (2013: Multispectral X ray Tomography)
- N. Chu (2013: Acoustic sources localization)
- Th. Boulay (2013: Non Cooperative Radar Target Recognition)
- R. Prenon (2013: Proteomic and Masse Spectrometry)
- Sh. Zhu (2012: SAR Imaging)
- D. Fall (2012: Emission Positon Tomography, Non Parametric Bayesian)
- D. Pougaza (2011: Copula and Tomography)
- H. Ayasso (2010: Optical Tomography, Variational Bayes)

Older Graduated PhD students:

- S. Fékih-Salem (2009: 3D X ray Tomography)
- N. Bali (2007: Hyperspectral imaging)
- O. Féron (2006: Microwave imaging)
- F. Humblot (2005: Super-resolution)
- M. Ichir (2005: Image separation in Wavelet domain)
- P. Brault (2005: Video segmentation using Wavelet domain)

Thanks to:

Older Graduated PhD students:

- H. Snoussi (2003: Sources separation)
- Ch. Soussen (2000: Geometrical Tomography)
- G. Montémont (2000: Detectors, Filtering)
- H. Carfantan (1998: Microwave imaging)
- S. Gautier (1996: Gamma ray imaging for NDT)
- M. Nikolova (1994: Piecewise Gaussian models and GNC)
- D. Prémel (1992: Eddy current imaging)

Post-Docs:

- J. Lapuyade (2011: Dimentionality Reduction and multivariate analysis)
- S. Su (2006: Color image separation)
- A. Mohammadpour (2004-2005: HyperSpectral image segmentation)

Colleagues:

- B. Duchêne & A. Joisel (L2S)& G. Perruson (Inverse scattering and Microwave Imaging)
- N. Gac (L2S) (GPU Implementation)
- Th. Rodet (L2S) (Computed Tomography)

Thanks to:

National Collaborators

- A. Vabre & S. Legoupil (CEA-LIST), (3D X ray Tomography)
- E. Barat (CEA-LIST) (Positon Emission Tomography, Non Parametric Bayesian)
- C. Comtat (SHFJ, CEA) (PET, Spatio-Temporal Brain activity)
- J. Picheral (SSE, Supélec) (Acoustic sources localization)
- D. Blacodon (ONERA) (Acoustic sources separation)
- J. Lagoutte (Thales Air Systems) (Non Cooperative Radar Target Recognition)
- P. Grangeat (LETI, CEA, Grenoble) (Proteomic and Masse Spectrometry)
- F. Lévi (CNRS-INSERM, Hopital Paul Brousse) (Biological rythms and Chronotherapy of Cancer)

International Collaborators

- K. Sauer (Notre Dame University, IN, USA) (Computed Tomography, Inverse problems)
- F. Marvasti (Sharif University), (Sparse signal processing)
- M. Aminghafari (Amir Kabir University) (Independent Components Analysis)
- A. Mohammadpour (AKU) (Statistical inference)
- Gh. Yari (Tehran Technological University) (Probability and Analysis)

References 1

- A. Mohammad-Djafari, "Bayesian approach with prior models which enforce sparsity in signal and image processing," EURASIP Journal on Advances in Signal Processing, vol. Special issue on Sparse Signal Processing, (2012).
- A. Mohammad-Djafari (Ed.) Problèmes inverses en imagerie et en vision (Vol. 1 et 2), Hermes-Lavoisier, Traité Signal et Image, IC2, 2009,
- A. Mohammad-Djafari (Ed.) Inverse Problems in Vision and 3D Tomography, ISTE, Wiley and sons, ISBN: 9781848211728, December 2009, Hardback, 480 pp.
- A. Mohammad-Djafari, Gauss-Markov-Potts Priors for Images in Computer Tomography Resulting to Joint Optimal Reconstruction and segmentation, International Journal of Tomography & Statistics 11: W09. 76-92, 2008.
- A Mohammad-Djafari, Super-Resolution : A short review, a new method based on hidden Markov modeling of HR image and future challenges, The Computer Journal doi:10,1093/comjnl/bxn005:, 2008.
- H. Ayasso and Ali Mohammad-Djafari Joint NDT Image Restoration and Segmentation using Gauss-Markov-Potts Prior Models and Variational Bayesian Computation, IEEE Trans. on Image Processing, TIP-04815-2009.R2, 2010.
- H. Ayasso, B. Duchene and A. Mohammad-Djafari, Bayesian Inversion for Optical Diffraction Tomography Journal of Modern Optics, 2008.
- N. Bali and A. Mohammad-Djafari, "Bayesian Approach With Hidden Markov Modeling and Mean Field Approximation for Hyperspectral Data Analysis," IEEE Trans. on Image Processing 17: 2. 217-225 Feb. (2008).
- H. Snoussi and J. Idier., "Bayesian blind separation of generalized hyperbolic processes in noisy and underdeterminate mixtures," IEEE Trans. on Signal Processing, 2006.

References 2

- O. Féron, B. Duchène and A. Mohammad-Djafari, Microwave imaging of inhomogeneous objects made of a finite number of dielectric and conductive materials from experimental data, Inverse Problems, 21(6):95-115, Dec 2005.
- M. Ichir and A. Mohammad-Djafari,
  Hidden markov models for blind source separation, IEEE Trans. on Signal Processing, 15(7):1887-1899, Jul 2006.
- F. Humblot and A. Mohammad-Djafari,
  Super-Resolution using Hidden Markov Model and Bayesian Detection Estimation Framework, EURASIP Journal on Applied Signal Processing, Special number on Super-Resolution Imaging: Analysis, Algorithms, and Applications:ID 36971, 16 pages, 2006.
- O. Féron and A. Mohammad-Djafari,
  Image fusion and joint segmentation using an MCMC algorithm, Journal of Electronic Imaging, 14(2):paper no. 023014, Apr 2005.
- H. Snoussi and A. Mohammad-Djafari,
  Fast joint separation and segmentation of mixed images, Journal of Electronic Imaging, 13(2):349-361, April 2004.
- A. Mohammad-Djafari, J.F. Giovannelli, G. Demoment and J. Idier,
  Regularization, maximum entropy and probabilistic methods in mass spectrometry data processing problems, Int. Journal of Mass Spectrometry, 215(1-3):175-193, April 2002.
- H. Snoussi and A. Mohammad-Djafari, "Estimation of Structured Gaussian Mixtures: The Inverse EM Algorithm," IEEE Trans. on Signal Processing 55: 7. 3185-3191 July (2007).
- N. Bali and A. Mohammad-Djafari, "A variational Bayesian Algorithm for BSS Problem with Hidden Gauss-Markov Models for the Sources," in: Independent Component Analysis and Signal Separation (ICA 2007) Edited by:M.E. Davies, Ch.J. James, S.A. Abdallah, M.D. Plumbley. 137-144 Springer (LNCS 4666) (2007).

References 3

- **N. Bali and A. Mohammad-Djafari, "Hierarchical Markovian Models for Joint Classification, Segmentation and Data Reduction of Hyperspectral Images" ESANN 2006, September 4-8, Belgium. (2006)**
- **M. Ichir and A. Mohammad-Djafari, "Hidden Markov models for wavelet-based blind source separation," IEEE Trans. on Image Processing 15: 7. 1887-1899 July (2005)**
- **S. Moussaoui, C. Carteret, D. Brie and A Mohammad-Djafari, "Bayesian analysis of spectral mixture data using Markov Chain Monte Carlo methods sampling," Chemometrics and Intelligent Laboratory Systems 81: 2. 137-148 (2005).**
- **H. Snoussi and A. Mohammad-Djafari, "Fast joint separation and segmentation of mixed images" Journal of Electronic Imaging 13: 2. 349-361 April (2004)**
- **H. Snoussi and A. Mohammad-Djafari, "Bayesian unsupervised learning for source separation with mixture of Gaussians prior," Journal of VLSI Signal Processing Systems 37: 2/3. 263-279 June/July (2004)**
- **F. Su and A. Mohammad-Djafari, "An Hierarchical Markov Random Field Model for Bayesian Blind Image Separation," 27-30 May 2008, Sanya, Hainan, China: International Congress on Image and Signal Processing (CISP 2008).**
- **N. Chu, J. Picheral and A. Mohammad-Djafari, "A robust super-resolution approach with sparsity constraint for near-field wideband acoustic imaging," IEEE International Symposium on Signal Processing and Information Technology pp 286–289, Bilbao, Spain, Dec14-17,2011**

Questions, Discussions, Open mathematical problems

- Sparsity representation, low rank matrix decomposition
  - Sparsity and positivity or other constraints
  - Group sparsity
  - Algorithmic and implementation issues for great dimensional applications (Big Data)
  - Joint estimation of Dictionary and coefficients
- Optimization of the KL divergence for Variational Bayesian Approximation
  - Convergency of alternate optimization
  - Other possible algorithms
- Properties of the obtained approximation
  - Does the moments of q's corresponds to the moments of p?
  - How about any other statistics: entropy, ...
- Other divergency or Distance measures?
- Using Sparsity as a prior in Inverse Problems
- Applications in Medical imaging, Non Destructive Testing (NDT) Industrial Imaging, Communication, Geophysical imaging, Radio Astronomy, ...